**IITD Machine Learning-01(chatbot)[INTP23-ML-1]**

**Day 7:**

On day 7, of IITD-AIA FSM internship, I applied beautiful soup as well as nlp.

**Learning:**

So, today I watched the beautiful soup and applied it side by side. So, in this website is been scrapped out and if we are searching using tags we ae getting right answers. But if we are searching for string or pattern then nothing is been returned or printed. Also, we are getting basic parsed file, I think user information is not there in the parsed data.

And, after attending yesterday’s doubt session I got to know various new concepts of NLP, which can be applied in deep learning. So, in order to learn new topics like word2vec, attention mechanism, first we should have initial pre-processing knowledge. So, I applied tokenization, stemming, lemmatization on random data and also bag of words now knowing the shortcoming of tokenization:

* sparse matrix (i.e., if features increase then matrix would become too hard to handle)
* out of vocabulary (i.e., if there are 4 words in vocab {girls, boys, food, good) and if new word like cat come then cat word would be ignored)
* semantic meaning is lost
* orders of words get changed, like there is features according to frequency.

We will learn word2vac that will improve the efficiency, now going step by step from pre-processing towards advantage technology. I will learn NLP in deep along with implementation.

Also, try to prepare the data simultaneously, which can be implemented into with NLP; applied with random data set.

**Work Done/learning Implemented:**

Applied beautiful soup on website data, and in order to learn word2vac I applied NLTK

**Is Progress As per Track?**

Yes, I started with coding part today and applied for basic chats for using word2vac.

**Issues Faced Today:** (None)

**Issues Closed Today:** (None)

**Highlights:**

So, today I watched the beautiful soup and applied it side by side.so, in this website is been scrapped out and if we are searching using tags we ae getting right answers. But if we are searching for string or pattern then nothing is been returned or printed. Also, we are getting basic parsed file, I think user information is not there in the parsed data.

Also, applied MLTK

**Concluding the day:**

After starting late due to extension in assignment of level 2, it’s a good start till now. So basically, I learnt NLTK module, where I studied various new things tokenizer, stemming, Regular expressions use of ‘re” library and how to extract data using regular expression and applied the same on random data

Also I applied website parser.

